

Concentration Inequalities in Locally Dependent Spaces

Daniel Paulin

*Department of Mathematics, National University of Singapore
10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore.
e-mail: paulindani@gmail.com*

Abstract: This paper studies concentration inequalities for functions of locally dependent random variables. We show that the usual definition of local dependence does not imply concentration for general Hamming Lipschitz functions. We define hypergraph dependence, which is a special case of local dependence, and show that it implies concentration if the maximal neighborhood size is small. We prove concentration in Hamming distance, Talagrand distance, and for self-bounding functions of a particular type under this dependence structure.

AMS 2000 subject classifications: Primary 60E15; secondary 60B20..

Keywords and phrases: Concentration inequality, local dependence, random matrices.

1. Introduction

Local dependence, when the variables only depend on those others which are in their neighborhood, has been one of the first examples of Stein's method, see ([Chen and Shao, 2004](#)) and the references therein.

The usual form of local dependence is the following (based on ([Chen, Goldstein and Shao, 2011](#)), Chapter 4.7.):

Definition 1. A group of random variables $\{X_i, i \in \mathcal{A}\}$ satisfies (LD1) if for each $i \in \mathcal{A}$ there exists $A_i \in [n]$ such that X_i and $X_{A_i^c}$ are independent.

An undirected graph $\mathcal{G} = ([n], \mathcal{E})$ is called a dependency graph of $\{X_i, i \in \mathcal{A}\}$ if each X_i can only depend on its neighbors in \mathcal{G} (i.e. it is independent of the complement of its neighbors). An example for such a graph \mathcal{G} is a graph with edge between i and j if $i \in A_j$ or $j \in A_i$ (i.e. one of them is in the neighborhood of the other).

We say that $\{X_i, i \in [n]\}$ satisfies (LD1, m) if there is a \mathcal{G} dependency graph that has maximum degree at most $m - 1$.

Let $\mathcal{G} = (V, E)$ be an undirected graph. The *chromatic number* of \mathcal{G} , $\chi(\mathcal{G})$ is the smallest positive integer k such that the vertices of \mathcal{G} can be colored with k colors with no edge between vertices of the same color.

([Janson, 2004](#)) shows concentration of sums under (LD1), and shows that Chernoff-Hoeffding and Bernstein inequalities also hold for sums of (LD1) dependent variables, with constants less than $\chi(\mathcal{G})$ times weaker than in the independent case.

The objective of this paper is to investigate whether this result holds for more general functions of (LD1) dependent variables. We could extend ([Janson, 2004](#)) to subadditive

functions. On the other hand, as the following counterexample shows, concentration does not hold for arbitrary Hamming-Lipschitz functions:

Example 1.1. For n even, let $X_1, \dots, X_{n/2}, X'_1, \dots, X'_{n/2}$ be random variables taking values 1 and -1. Let $X_1, \dots, X_{n/2}$ be i.i.d. with $P(X_i = 1) = P(X_i = -1) = 1/2$. Let Q be an independent random variable with $P(Q = 1) = P(Q = -1) = 1/2$. Define $X'_i = Q \cdot X_i$, $1 \leq i \leq n/2$.

For $\{X_1, \dots, X_{n/2}, X'_1, \dots, X'_{n/2}\}$ as defined such satisfy (LD1,2). Define the function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ as $g(a, b) = a \cdot b/2$, then $g(1, 1) = g(-1, -1) = 1/2$ and $g(1, -1) = g(-1, 1) = -1/2$. Define $f(X_1, \dots, X_{n/2}, X'_1, \dots, X'_{n/2}) = \sum_{i=1}^{n/2} g(X_i, X'_i)$, then this f is 1-Hamming Lipschitz (depending on n variables). On the other hand, for the distribution we gave to X_i and X'_i , we have $g(X_i, X'_i) = Q$, so $f(X_1, \dots, X_{n/2}, X'_1, \dots, X'_{n/2}) = nQ/2$, taking values $n/2$ and $-n/2$ with probability $1/2$.

We have looked for examples in the literature about (LD1) dependent random variables, and most of them were defined as functions of independent random variables. For such cases, as we will show, concentration inequalities hold for general functions.

1.1. Main definitions

We will use the fractional chromatic number:

Definition. Let $\mathcal{G} = (V, E)$ be an undirected graph. The fractional chromatic number of \mathcal{G} , $\chi^*(\mathcal{G})$ is the smallest positive real k for which there exists a probability distribution over the independent sets of \mathcal{G} such that for each vertex v , given an independent set S drawn from the distribution,

$$\Pr(v \in S) \geq \frac{1}{k}.$$

The independent sets of \mathcal{G} here mean all the subsets of the vertices of \mathcal{G} that contain no edges between them.

(Janson, 2004) introduces these:

Definition. Given \mathcal{A} and $\{X_\alpha\}$, $\alpha \in \mathcal{A}$, we make the following definitions:

- A subset \mathcal{A}' of \mathcal{A} is independent if the corresponding random variables $\{X_\alpha\}_{\alpha \in \mathcal{A}'}$ are independent.
- A family $\{\mathcal{A}_j\}_j$ of subsets of \mathcal{A} is a cover of \mathcal{A} if $\cup_j \mathcal{A}_j = \mathcal{A}$.
- A family $\{(\mathcal{A}_j, w_j)\}_j$ of pairs (\mathcal{A}_j, w_j) where $\mathcal{A}_j \subset \mathcal{A}$ and $w_j \in [0, 1]$ is a fractional cover of \mathcal{A} if $\sum_{j: \alpha \in \mathcal{A}_j} w_j \geq 1$ for each $\alpha \in \mathcal{A}$.
- A (fractional) cover is proper if each set \mathcal{A}_j is independent.
- $\chi(\mathcal{A})$ is the size of the smallest proper cover of \mathcal{A} , i.e. the smallest m such that \mathcal{A} is the union of m independent subsets
- $\chi^*(\mathcal{A})$ is the minimum of $\sum_j w_j$ over all proper fractional covers $\{(\mathcal{A}_j, w_j)\}_j$.
- We say that a fractional cover $\{(\mathcal{A}_j, w_j)\}_j$ is exact if $\sum_j w_j \mathbb{1}_{\mathcal{A}_j} = \mathbb{1}_{\mathcal{A}}$.

It is shown in (Janson, 2004) that for (LD1, m) with dependency graph \mathcal{G} ,

$$\chi^*(\mathcal{A}) \leq \chi^*(\mathcal{G}) \leq \chi(\mathcal{G}) \leq m.$$

Lemma 3.2 of (Janson, 2004) shows that we can make an exact fractional cover from any fractional cover without changing $\sum_j w_j$, thus we can restrict our attention to exact fractional covers.

The first result for sums is the following:

Theorem 1.1 (Theorem 2.1 of (Janson, 2004)). *Suppose that $\{X_i\}_{i \in \mathcal{A}}$ satisfies (LD1), $a_i \leq X_i \leq b_i$ for $i \in \mathcal{A}$ and some real numbers a_i and b_i . Then, for $t > 0$, Denote $S := \sum_{i \in \mathcal{A}} X_i$, then for $t > 0$,*

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp \left(-2 \frac{t^2}{\chi^*(\mathcal{A}) \sum_{i \in \mathcal{A}} (b_i - a_i)^2} \right). \quad (1.1)$$

The same estimate holds for $\mathbb{P}(X \leq \mathbb{E}X - t)$.

Further Bernstein-type results are proven in (Janson, 2004), which take into account the variance of X_i , and thus give better bounds than (1.2) for sums of random variables with small variances, an example is the following:

Theorem 1.2 (Theorem 2.3 of (Janson, 2004)). *Suppose that $\{X_i\}_{i \in \mathcal{A}}$ satisfies (LD1), $X_i - \mathbb{E}X_i \leq b$ for some $b > 0$, and all $i \in \mathcal{A}$. Then, for $t > 0$, Denote $S := \sum_{i \in \mathcal{A}} \text{Var}(X_i)$, then for $t > 0$,*

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp \left(-\frac{8t^2}{25\chi^*(\mathcal{A})(S + bt/3)} \right). \quad (1.2)$$

The main idea of the proofs in (Janson, 2004) is to separate $\sum_{i \in \mathcal{A}} X_i$ into sums of independent random variables, and use the concentration properties of such sums.

Our first result, Theorem 2.1, gives an upper tail bound for subadditive functions. The proof is based on the same idea as (Janson, 2004). An application is given: an estimate for the upper tail of the norm of random matrix with locally dependent entries.

Example 1.1 made us look for other, stronger definitions of local dependence, that are sufficient for concentration for large class of functions ((GD) is based on (Chen, Goldstein and Shao, 2011)).

Definition 2 (GD). $\{X_i, i \in \mathcal{A}\}$ satisfies graphical dependence if we can define a graph $\mathcal{G} = (\mathcal{A}, \mathcal{E})$ such that for any pair of disjoint sets Γ_1, Γ_2 in \mathcal{A} such that there is no edge in \mathcal{E} that has one endpoint in Γ_1 and another in Γ_2 , the sets of random variables X_{Γ_1} and X_{Γ_2} are independent. In this case \mathcal{G} is called the dependency graph. We say $\{X_i, i \in [n]\}$ satisfies (GD, m) if \mathcal{G} has maximum degree at most $m - 1$.

Definition 3 (HD). Let $\{Y_i, i \in [N]\}$ be a set of independent random variables taking values in $\Sigma = \Sigma_1 \times \dots \times \Sigma_N$, and for each $i \in [n]$, let S_i be subsets of $[N]$, and $X_i : Y_{S_i} \rightarrow \Lambda_i$ be random variables depending on Y_{S_i} . For each $j \in [N]$, let $R_j := \{i \in [n] \text{ s.t. } j \in S_i\}$ (i.e. R_j the set of X_i depending of Y_j , and S_i is the set of Y_j that X_i depends on).

We say that $\{X_i, i \in [n]\}$ satisfies (HD, k, l) if $|S_i| \leq k$ and $|R_j| \leq l$ for every $i \in [n], j \in [N]$. Let $\mathcal{G} = ([n], \mathcal{E})$ be an undirected graph with an edge between i and j if X_i and X_j depend on some common Y_k (i.e. if $S_i \cap S_j \neq \emptyset$). If \mathcal{G} has maximum degree at most $m - 1$, then we say $\{X_i : i \in [n]\}$ satisfies (HD, m).

A relation between (HD, k, l) and (HD, m) is given by the following lemma (the proof is given in Section 4):

Lemma 1.1. *Suppose that $\{X_i : i \in [n]\}$ satisfies (HD, m) for some $\{Y_i, i \in [N]\}$, then we can define $\{Y'_i, i \in [N']\}$ such that $\{X_i : i \in [n]\}$ satisfies (HD, m, m) for $\{Y'_i, i \in [N']\}$.*

Example 1.2 (m -dependence). *A simple example to illustrate the difference between (HD, k, l) and (HD, m) is the following: let Y_1, \dots, Y_n be independent random variables, and*

$$X_1 := f_1(Y_1, \dots, Y_m), X_2 := f_2(Y_2, \dots, Y_{m+1}), \dots, X_n := f_n(Y_n, Y_1, \dots, Y_{m-1}).$$

Then one can easily prove (by breaking Y into groups of size m) that X_1, \dots, X_n satisfy $(HD, 2, 2m - 1)$ and $(HD, 2m - 1)$.

Example 1.3 (Triangles in Erdős-Rényi graph). *Let $G(n, p)$ be an Erdős-Rényi graph, with edges $(X_{ij})_{1 \leq i < j \leq n}$ (i.e. X_{ij} is the indicator function of the edge between i and j), and denote by $(T_{ijk})_{1 \leq i < j < k \leq n}$ the indicator functions of the triangles between vertices i, j, k . Then one can easily see that $(T_{ijk})_{1 \leq i < j < k \leq n}$ satisfies $(HD, n - 2, 3)$ and $(HD, 3(n - 3))$.*

It is an easy exercise to prove that

$$(HD, m) \Rightarrow (GD, m) \Rightarrow (LD1, m).$$

The reverse implications are false in general.

$(LD1, m)$ does not imply (GD, m) , as we can see from Example 1.1.

(GD, m) does not imply (HD, m) , we can see this from the example in (Burton, Goulet and Meester, 1993) where they construct a one - dependent sequence (future independent of past) which only satisfies (GD, m) (the existence of such a sequence was an open question for many years).

It remains an open question whether (GD, m) implies concentration inequalities for general functions. At the moment, we do not know of practical applications that satisfy (GD, m) , but not (HD, m) .

1.2. Self-bounding and α -self-bounding functions

Self bounding functions were introduced in (Boucheron, Lugosi and Massart, 2000), and found many applications. In (Boucheron, Lugosi and Massart, 2009), the authors introduce (a, b) self-bounding and weakly (a, b) self-bounding functions.

For independent random variables, (a, b) self-bounding functions are a large class of functions, that contain Hamming Lipschitz functions, configuration functions, suprema of positive valued empirical processes. They also imply Talagrand's convex distance inequality.

In (Paulin, 2012a), we have defined a stronger condition, α -(a, b) self-bounding and weakly α -(a, b) self-bounding functions, and shown that such functions satisfy concentration inequalities under some dependence condition. As we are going to see, they also satisfy concentration inequalities under the (HD, k, l) dependence condition.

The following definitions of self-bounding functions are from (Boucheron, Lugosi and Massart, 2009) (we made a slight generalization, they had $\Lambda_1 = \dots = \Lambda_n = \mathcal{X}$).

For $1 \leq i \leq n$, and $x \in \Lambda$, let $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, and let $\Lambda_{-i} := \Lambda_1 \times \dots \times \Lambda_{i-1} \times \Lambda_{i+1} \times \dots \times \Lambda_n$.

Definition 4. *A function $g : \Lambda \rightarrow \mathbb{R}$ is called (a, b) -self-bounding for some $a, b \geq 0$ if there are functions $g_i : \Lambda_{-i} \rightarrow \mathbb{R}$ such that for all $i = 1, \dots, n$ and all $x \in \Lambda$,*

1. $0 \leq g(x) - g_i(x_{-i}) \leq 1$, and
2. $\sum_{i=1}^n (g(x) - g_i(x_{-i})) \leq ag(x) + b$.

A function $g : \Lambda \rightarrow \mathbb{R}$ is called weakly (a, b) -self-bounding if there are functions $g_i : \Lambda_{-i} \rightarrow \mathbb{R}$ such that for all $x \in \Lambda$,

$$\sum_{i=1}^n (g(x) - g_i(x_{-i}))^2 \leq ag(x) + b.$$

Remark 1.1. If g is (a, b) -self-bounding, then it is also (a, b) -weakly-self-bounding. If g is (a, b) -self-bounding for some g_i , then it is also (a, b) -self-bounding for

$$g_i(x_{-i}) := \inf_{x'_i \in \Lambda_i} f(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n). \quad (1.3)$$

If g is weakly (a, b) -self-bounding, then in this paper we will also assume that $g_i(x_{-i}) \leq g(x)$ for all $x \in \Lambda$, and in this case, we can choose g_i as in (1.3). In the rest of this paper, we assume that g_i is chosen as (1.3).

For these functions, the following concentration inequalities hold ((Boucheron, Lugosi and Massart, 2009) supposed that $\Lambda_1 = \dots = \Lambda_n = \mathcal{X}$, but the same results trivially hold for this case):

Theorem 1.3. ((Boucheron, Lugosi and Massart, 2009)) Let

$$X := (X_1, \dots, X_n)$$

be a vector of independent random variables, taking values in Λ and let $f : \Lambda \rightarrow \mathbb{R}$ be a non-negative measurable function such that $Z = f(X)$ has finite mean. For $a, b \geq 0$, define $c = (3a - 1)/6$. If f is (a, b) -self-bounding, then for all $\lambda \geq 0$,

$$\log \mathbb{E} [e^{\lambda(Z - \mathbb{E}Z)}] \leq \frac{(a\mathbb{E}Z + b)\lambda^2}{2(1 - c_+\lambda)}$$

For all $t > 0$,

$$\mathbb{P}\{Z \geq \mathbb{E}Z + t\} \leq \exp \left(-\frac{t^2}{2(a\mathbb{E}Z + b + c_+t)} \right).$$

If f is weakly (a, b) -self-bounding and for all $i \leq n$, all $x \in \Lambda$, $f_i(x^{(i)}) \leq f(x)$, then for all $0 \leq \lambda \leq 2/a$,

$$\log \mathbb{E} [e^{\lambda(Z - \mathbb{E}Z)}] \leq \frac{(a\mathbb{E}Z + b)\lambda^2}{2(1 - a\lambda/2)}$$

and for all $t > 0$,

$$\mathbb{P}\{Z \geq \mathbb{E}Z + t\} \leq \exp \left(-\frac{t^2}{2(a\mathbb{E}Z + b + at/2)} \right).$$

If f is weakly (a, b) -self-bounding and $f(x) - f_i(x^{(i)}) \leq 1$ for each $i \leq n$ and $x \in \Lambda$, then for $0 \leq t \leq \mathbb{E}Z$,

$$\mathbb{P}\{Z \leq \mathbb{E}Z - t\} \leq \exp \left(-\frac{t^2}{2(a\mathbb{E}Z + b + c_-t)} \right).$$

We define α -self-bounding functions as in (Paulin, 2012a):

Definition 5. Let $\Omega = \Omega_1 \times \dots \times \Omega_n$. Let $a, b \geq 0$.

1. We say that $f : \Omega \rightarrow \mathbb{R}$ is α -(a, b) self-bounding if there is $\alpha : \Lambda \rightarrow \mathbb{R}_+^n$ such that
 - (a) $f(x) - f(y) \leq \sum_{i \leq n} \alpha_i(x) \mathbb{1}[x_i \neq y_i]$ for every $x, y \in \Omega$.
 - (b) $\alpha_i(x) \leq 1$ for every $i \leq n, x \in \Omega$.
 - (c) $\sum_{i \leq n} \alpha_i(x) \leq af(x) + b$.
2. We say that $f : \Omega \rightarrow \mathbb{R}$ is weakly α -(a, b) self-bounding if there is $\alpha : \Lambda \rightarrow \mathbb{R}_+^n$ such that
 - (a) $f(x) - f(y) \leq \sum_{i \leq n} \alpha_i(x) \mathbb{1}[x_i \neq y_i]$ for every $x, y \in \Omega$.
 - (b) $\sum_{i \leq n} \alpha_i(x)^2 \leq af(x) + b$.

Remark 1.2. It is easy to see that α -(a, b) self-bounding functions are also weakly α -(a, b) self-bounding.

Remark 1.3. The following relations hold:

$$\begin{array}{ccc}
 (a, b)\text{-self-bounding} & \Rightarrow & \text{weakly } (a, b)\text{-self-bounding} \\
 \uparrow & & \uparrow \\
 \alpha\text{-(}a, b\text{)-self-bounding} & \Rightarrow & \text{weakly } \alpha\text{-(}a, b\text{)-self-bounding}
 \end{array}$$

The reverse implications are false in general.

2. Results

The following theorem bounds the moment generating function of subadditive functions of (LD1) variables.

Theorem 2.1. Suppose that $\Lambda := \mathbb{R}^n$ (or \mathbb{R}^n), and $f : \Lambda \rightarrow \mathbb{R}$ is a subadditive function, i.e. for any $x, y \in \Lambda$, $f(x + y) \leq f(x) + f(y)$.

Let $\mathcal{A} = [n]$, and suppose that $\{X_i\}_{i \in \mathcal{A}}$ satisfy (LD1). Then

- If $\{\mathcal{A}_j\}_j$ is one of the smallest proper covers of \mathcal{A} , having $\chi(\mathcal{A})$ elements, then for $\theta > 0$,

$$\mathbb{E} \left(e^{\theta f(X_1, \dots, X_n)} \right) \leq \frac{1}{\chi(\mathcal{A})} \sum_{j=1}^{\chi(\mathcal{A})} \mathbb{E} \left(e^{\theta \chi(\mathcal{A}) f(X_{\mathcal{A}_j})} \right), \quad (2.1)$$

here $X_{\mathcal{A}_j} \in \Lambda$ is defined by replacing all the components of X with zeros outside of \mathcal{A}_j .

- Suppose that f also satisfies $f(cx) \leq cf(x)$ for every $0 \leq c \leq 1$, $x \in \Lambda$. Let $\{(A_j, w_j)\}_j$ be an exact fractional cover with $\sum_j w_j = \chi^*(\mathcal{A})$. Then for every $\theta > 0$,

$$\mathbb{E} \left(e^{\theta f(X_1, \dots, X_n)} \right) \leq \frac{1}{\chi^*(\mathcal{A})} \sum_j w_j \mathbb{E} \left(e^{\theta \chi^*(\mathcal{A}) f(X_{A_j})} \right). \quad (2.2)$$

Our next theorem is about (HD, k, l) dependent random variables.

Theorem 2.2. *Let $X = (X_1, \dots, X_n)$ be a Λ valued random vector, satisfying (HD, k, l) for some $Y = (Y_1, \dots, Y_N)$.*

- *If $f : \Lambda \rightarrow \mathbb{R}$ is α -(a, b)-self-bounding for some $a, b \geq 0$, then there is a function g such that $f(X) = g(Y)$ almost surely, and the function h defined as $h(y) := g(y)/l$ is $(ka, \frac{k}{l}b)$ -self-bounding.*
- *If $f : \Lambda \rightarrow \mathbb{R}$ is weakly α -(a, b)-self-bounding for some $a, b > 0$, then there is a function g such that $f(X) = g(Y)$ almost surely, and the function h defined as $h(y) := g(y)/l$ is weakly $(ka, \frac{k}{l}b)$ -self-bounding, and satisfies $|h(z) - h(z')| \leq 1$ for any z, z' differing only in one coordinate.*

Let $\Lambda := \Lambda_1 \times \dots \times \Lambda_n$, we say that a function $f : \Lambda \rightarrow \mathbb{R}$ is c -weighted Hamming Lipschitz for some $c \in \mathbb{R}_+^n$ if for any $x, y \in \Lambda$ only differing in coordinate i , $|f(x) - f(y)| \leq c_i$.

Corollary 2.1. *Suppose that $X = \{X_i, i \in [n]\}$ satisfies (HD, l, k) , $X \in \Lambda$, then for any c -weighted Hamming Lipschitz $f : \Lambda \rightarrow \mathbb{R}$, we have for every $\lambda > 0$,*

$$\mathbb{E}(\exp(\lambda[f(X) - \mathbb{E}f(X)])) \leq \exp\left(\frac{\lambda^2 kl \sum_{i=1}^n c_i^2}{8}\right), \quad (2.3)$$

and thus for every $t \geq 0$

$$\mathbb{P}(f(X) - \mathbb{E}f(X) \geq t), \mathbb{P}(f(X) - \mathbb{E}f(X) \leq -t) \leq \exp\left(\frac{-2t^2}{kl \sum_{i=1}^n c_i^2}\right). \quad (2.4)$$

Corollary 2.2. *Suppose that $X = \{X_i, i \in [n]\}$ satisfies (HD, l, k) , then a version of Talagrand's convex distance inequality holds:*

$$\mathbb{E}\left(\exp\left(\frac{1}{10kl}d_T^2(X, S)\right)\right) \leq \frac{1}{\mathbb{P}(X \in S)}. \quad (2.5)$$

and as a consequence,

$$\mathbb{P}(X \in S)\mathbb{P}(X \in \overline{S_t}) \leq \exp(-t^2/(10kl)). \quad (2.6)$$

3. Applications

Random matrix models with dependent entries have been considered by several authors, see for example (Anderson and Zeitouni, 2008). A model where the entries are (LD1) type, and satisfy some additional condition, appears in (Schenker and Schulz-Baldes, 2005), and then was further developed in (Hofmann-Credner and Stolz, 2008). These results show asymptotic convergence of the eigenvalue distribution to circular law or the singular value distribution to Marchenko-Pastur law. In this paper, we prove some non-asymptotic results. In our first example, we show concentration for the upper tail of the norm of a random matrix with (LD1) entries:

3.1. Norm of a random matrix with (LD1) entries

Theorem 3.1. *Let $M := (X_{i,j})_{1 \leq i,j \leq n}$ be a Hermitian matrix with entries bounded by K in absolute value, and the upper diagonal entries $(X_{i,j})_{1 \leq i \leq j \leq n}$ satisfying (LD1) with neighborhoods $\mathcal{A} = \{\mathcal{A}_{i,j}\}_{1 \leq i \leq j \leq n}$. Then*

$$\mathbb{P}(\|M\| \geq 3C\chi^*(\mathcal{A})K\sqrt{n} + t) \leq \exp\left(-\frac{t^2}{32\chi^*(\mathcal{A})^2K^2}\right) \quad (3.1)$$

here C is the universal constant in (Latała, 2005).

If $M := (X_{i,j})_{1 \leq i \leq n, j \leq N}$ is a complex valued matrix, with entries satisfying (LD1) with neighborhoods $\mathcal{A} = \{\mathcal{A}_{i,j}\}_{1 \leq i \leq j \leq n}$, then

$$\mathbb{P}(\|M\| \geq C\chi^*(\mathcal{A})K(\sqrt{n} + \sqrt{N} + \sqrt[4]{nN}) + t) \leq \exp\left(-\frac{t^2}{8\chi^*(\mathcal{A})^2K^2}\right) \quad (3.2)$$

3.2. Eigenvalues of a random matrix with (HD,k,l) entries

The following theorem is a generalization of Theorem 1 of (Alon, Krivelevich and Vu, 2002) to this setting.

Theorem 3.2. *Let M be a real valued random symmetric matrix with entries bounded by 1, and the upper diagonal entries satisfying (HD,k,l). Let $\lambda_1(M) \geq \dots \geq \lambda_n(M)$ be the eigenvalues of M in decreasing order. For every positive integer $1 \leq s \leq n$, the probability that $\lambda_s(M)$ deviates from its median by more than t is at most $4e^{-t^2/(80s^2 \cdot kl)}$. The same estimate holds for the probability that λ_{n-s+1} deviates from its median by more than t .*

Remark 3.1. We leave it to the reader as an exercise to adapt the proof of (Meckes, 2004), Theorem 2 to this setting, and reduce the s^2 in the exponent to s .

Remark 3.2. The correct range of concentration of the eigenvalues of a Wigner matrix M_n is $O\left(\sqrt{\frac{\log(n)}{n}}\right)$ in the bulk, and $O(n^{-1/6})$ on the edge, as it is shown in (Dallaporta, 2012).

4. Proofs

Proof of Lemma 1.1. Given Y_1, \dots, Y_N and S_1, \dots, S_n , let us define S'_1, \dots, S'_n the following way: for every $1 \leq i \leq N$, $i \in S'_j$ if and only if $1 \leq j \leq n$ is the smallest index such that $i \in S_j$. Now let $Y'_1 := Y_{S'_1}, \dots, Y'_n := Y_{S'_n}$, then the reader can easily verify that X_1, \dots, X_n satisfies (HD,m,m) for $\{Y'_i, i \in [n]\}$. \square

Proof of Theorem 2.1. Part 1 is implied by the subadditivity of f , and the convexity of the exponential function. For $\theta > 0$,

$$\begin{aligned} e^{\theta f(X)} &\leq e^{\theta(f(X_{\mathcal{A}_1}) + \dots + f(X_{\mathcal{A}_{\chi(\mathcal{A})}}))} = e^{\theta \chi(\mathcal{A}) \sum_{i=1}^{\chi(\mathcal{A})} \frac{1}{\chi(\mathcal{A})} f(X_{\mathcal{A}_i})} \\ &\leq \frac{1}{\chi(\mathcal{A})} \sum_{i=1}^{\chi(\mathcal{A})} e^{\theta \chi(\mathcal{A}) f(X_{\mathcal{A}_i})} \end{aligned}$$

Taking expectations gives the result. Part 2 is similar: for $\theta > 0$,

$$\begin{aligned} e^{\theta f(X)} &= e^{\theta f(\sum_i w_i X_{\mathcal{A}_i})} \leq e^{\theta \sum_i f(w_i X_{\mathcal{A}_i})} \leq e^{\theta \sum_i w_i f(X_{\mathcal{A}_i})} \\ &\leq e^{\theta \chi^*(\mathcal{A}) \sum_i \frac{w_i}{\chi^*(\mathcal{A})} f(X_{\mathcal{A}_i})} \leq \sum_i \frac{w_i}{\chi^*(\mathcal{A})} e^{\theta \chi^*(\mathcal{A}) f(X_{\mathcal{A}_i})}. \end{aligned}$$

Taking expectations gives the result. We have used the $f(cx) \leq cf(x)$ condition (for $0 \leq c \leq 1$, since $0 \leq w_i \leq 1$ for exact covers). \square

Proof of Theorem 2.2. We are only going to prove the first part (concerning α -self-bounding functions), the second part is similar.

The existence of g such that $f(X) = g(Y)$ is trivial, since X is a function of Y . For each $1 \leq i \leq N$, $y \in \Sigma$, let

$$h_i(y_{-i}) := \inf_{y'_i} h(y_1, \dots, y_{i-1}, y'_i, y_{i+1}, \dots, y_n).$$

Using the fact that $f(X)$ is $\alpha(a, b)$ self-bounding, we can write

$$\begin{aligned} \sum_{i=1}^N h(y) - h_i(y_{-i}) &\leq \frac{1}{l} \sum_{i=1}^N \sum_{j \in R_i} \alpha_j(x(y)) \\ &\leq \frac{1}{l} k \sum_{j=1}^n \alpha_j(x(y)) \leq \frac{k}{l} (af(x(y)) + b) \leq kh(y) + \frac{k}{l} b, \end{aligned}$$

and thus the result follows. \square

Proof of Corollary 2.1. This follows by a 4 times worse constant from the fact that f is weakly $\alpha(0, \sum_{i=1}^n c_i^2)$ self-bounding. We can get this better constant by writing $f(X) = g(Y)$, and directly applying Mcdiarmid's bounded differences inequality to the independent variables Y_1, \dots, Y_N . \square

Proof of Corollary 2.2. The proof is similar to the proof of Corollary 1 of (Boucheron, Lugosi and Massart, 2009). By Lemma 3.2 of (Paulin, 2012b), we know that $d_T^2(x, S)$ is weakly $\alpha(4, 0)$ self-bounding. By Theorem 2.2 we have a function g with $g(Y) = d_T^2(X, S)$, and $h(y) = g(y)/l$ is $(4k, 0)$ self-bounding.

Thus, by Theorem 1.3, we have that for $0 \leq \lambda \leq \frac{1}{2kl}$,

$$\begin{aligned} \log \mathbb{E}(e^{\lambda(d_T^2(X, S) - \mathbb{E}(d_T^2(X, S)))}) &= \log \mathbb{E}(e^{\lambda(h(Y) - \mathbb{E}(h(Y)))}) \\ &\leq \frac{4kl\lambda^2 \mathbb{E}(d_T^2(X, S))}{2(1 - 2kl\lambda)}, \end{aligned}$$

thus

$$\log \mathbb{E}(e^{\lambda d_T^2(X, S)}) \leq \lambda \mathbb{E}(d_T^2(X, S)) + \frac{4kl\lambda^2}{2(1 - 2kl\lambda)} \mathbb{E}(d_T^2(X, S)). \quad (4.1)$$

Again by Theorem 1.3,

$$\begin{aligned} \log \mathbb{P}(X \in S) &= \log \mathbb{P}(d_T^2(X, S) - \mathbb{E}d_T^2(X, S) \leq -\mathbb{E}d_T^2(X, S)) \\ &= \log \mathbb{P}(h(Y) - \mathbb{E}h(Y) \leq -\mathbb{E}h(Y)) \leq -\frac{\mathbb{E}h(Y)}{2 \cdot 4k} = -\frac{\mathbb{E}d_T^2(X, S)}{8kl}. \end{aligned}$$

By adding this to (4.1) for $\lambda = \frac{1}{10kl}$, we get the result. \square

Before proving Theorem 3.1, we introduce some results that we are going to use:

Theorem 4.1 (Theorem 2 of (Latała, 2005)). *For any finite matrix of independent mean zero r.v.'s X_{ij} we have*

$$\mathbb{E}||X_{ij}|| \leq C \left(\max_i \sqrt{\sum_j \mathbb{E}X_{ij}^2} + \max_j \sqrt{\sum_i \mathbb{E}X_{ij}^2} + \sqrt[4]{\sum_{ij} \mathbb{E}X_{ij}^4} \right)$$

Proposition 4.1. *Let $A_{i,j}$ be a symmetric real valued matrix with entries bounded by K . Let $\lambda_1(A_{i,j})$ be its largest eigenvalue. If we look at λ_1 as a function of only $A_{1 \leq i \leq j \leq n}$, then λ_1 is weakly $(0, 16K^2)$ self-bounding.*

Proof. This is a reformulation of Example of page 42-43 of (Lugosi, 2005) (but they made a mistake by treating all the elements of the symmetric matrix as independent random variables, so the correct constant is 4 times worse).

For some $v \in \mathbb{R}^n$,

$$\lambda_1(A) = \sup_{u \in \mathbb{R}^n: ||u||=1} u^t A u = v^t A v = \sum_{i,j} v_i v_j A_{ij},$$

and for A' created by replacing $A_{i,j}$ and $A_{j,i}$ with $A'_{i,j}$, we have

$$\lambda_1(A) - \lambda_1(A') \leq 2v_i v_j (A_{i,j} - A'_{i,j}) \leq 4K |v_i| |v_j|,$$

so the result follows by

$$\sum_{i,j} (4K |v_i| |v_j|)^2 \leq 16K^2 \left(\sum_i |v_i|^2 \right) \left(\sum_j |v_j|^2 \right) \leq 16K^2.$$

\square

In the following proposition, we prove a similar result for largest singular value:

Proposition 4.2. *Let $A := (A_{i,j})$ be an $n \times N$ sized complex valued matrix with entries bounded by K in absolute value. Let $s_1(A)$ be its largest singular value (which is equal to its operator norm). Then s_1 is weakly $(0, 4K^2)$ self-bounding.*

If $A := (A_{i,j})$ is a Hermitian $n \times n$ matrix, then s_1 , as the function of only $A_{1 \leq i \leq j \leq n}$, is weakly $(0, 16K^2)$ self-bounding.

Proof. We can write, for some U, V complex valued unit vectors, that

$$\begin{aligned} s_1(A) &= \sup_{u,v \in \mathbb{R}^n: ||u||, ||v||=1} \operatorname{Re}[u^* A v] = \sup_{u,v: ||u||, ||v||=1} \sum_{ij} \operatorname{Re}[u_i^* v_j A_{ij}] \\ &= \sum_{ij} \operatorname{Re}[U_i^* V_j A_{ij}]. \end{aligned}$$

Let's denote by A' the matrix that we get by changing A_{ij} to A'_{ij} , then one can easily see that

$$s_1(A) - s_1(A') \leq \operatorname{Re}[U_i^* V_j A_{ij}] - \operatorname{Re}[U_i^* V_j A'_{ij}] \leq 2K |U_i| |V_j|,$$

and thus the first statement of the proposition is implied by

$$\sum_{ij} (2K|U_i||V_j|)^2 \leq 4K^2 \left(\sum_{1 \leq i \leq n} |U_i|^2 \right) \left(\sum_{1 \leq j \leq N} |V_j|^2 \right) \leq 4K^2.$$

The proof of the last statement is similar, and is left to the reader. \square

Proof of Theorem 3.1. In the Hermitian case, let $\Lambda = \mathbb{R}^{n(n-1)/2}$. Let $f : \Lambda \rightarrow \mathbb{R}$ be the norm of the Hermitian matrix with upper diagonal elements as argument. Let $\mathcal{A} = [n(n-1)/2]$, and let (\mathcal{A}_i, w_i) be an exact cover for \mathcal{A} with $\sum_i w_i = \chi^*(\mathcal{A})$.

Then $f(cx) = cf(x)$ for $0 \leq c \leq 1$, so by the second part of Theorem 2.1,

$$\mathbb{E}(e^{\theta \|M\|}) \leq \frac{1}{\chi^*(\mathcal{A})} \sum_j w_j \mathbb{E}(e^{\theta \chi^*(\mathcal{A}) \|M_{\mathcal{A}_j}\|}).$$

Now for each j , $M_{\mathcal{A}_j}$ is a symmetric matrix with independent entries, and thus, by the second statement of Proposition 4.2, we know that $\|M_{\mathcal{A}_j}\|$ is weakly $(0, 16K^2)$ self-bounding as a function of its upper diagonal entries. This means that by Theorem 1.3, we have for every $\lambda > 0$, for every j ,

$$\mathbb{E}(e^{\lambda \|M_{\mathcal{A}_j}\|}) \leq e^{\lambda \mathbb{E}\|M_{\mathcal{A}_j}\|} \cdot e^{8K^2 \lambda^2},$$

which, by setting $\lambda := \chi^*(\mathcal{A})\theta$, implies that

$$\mathbb{E}(e^{\theta \|M\|}) \leq e^{8K^2 \chi^*(\mathcal{A})^2 \theta^2} \frac{1}{\chi^*(\mathcal{A})} \sum_j w_j e^{\theta \chi^*(\mathcal{A}) \mathbb{E}\|M_{\mathcal{A}_j}\|}. \quad (4.2)$$

Now, Theorem 4.1, combined with the boundedness assumption, implies that

$$\mathbb{E}\|M_{\mathcal{A}_j}\| \leq 3C\sqrt{n},$$

thus by Markov's inequality, we get

$$\mathbb{P}(\|M\| \geq 3C\chi^*(\mathcal{A})K\sqrt{n} + t) \leq \exp(8K^2 \chi^*(\mathcal{A})^2 \theta^2 - \theta t),$$

and taking $\theta = \frac{t}{16\chi^*(\mathcal{A})^2}$ gives the bound.

The proof of the rectangular case is similar. \square

Proof of Theorem 3.2. This is just a simple adaptation of the argument of (Alon, Krivelevich and Vu, 2002). The version of Talagrand's inequality for (HD, k, l) dependence, which follows from Corollary 2.2, is of the form

$$\Pr[\mathcal{A}] \Pr[\mathcal{B}] \leq e^{-t^2/(10kl)},$$

and thus the constant the exponent becomes $80kl$ (instead of 32). \square

Acknowledgements

The author thanks Doma Szász and Mogyi Tóth for infecting him with their enthusiasm of probability. He thanks his thesis supervisors, Louis Chen and Adrian Röllin, for the opportunity to study in Singapore, and their useful advices. Finally, many thanks to my brother, Roland Paulin, for the enlightening discussions.

References

- ALON, N., KRIVELEVICH, M. and VU, V. H. (2002). On the concentration of eigenvalues of random symmetric matrices. *Israel J. Math.* **131** 259–267. . [MR1942311 \(2003i:15020\)](#)
- ANDERSON, G. W. and ZEITOUNI, O. (2008). A law of large numbers for finite-range dependent random matrices. *Comm. Pure Appl. Math.* **61** 1118–1154. . [MR2417889 \(2010a:60091\)](#)
- BOUCHERON, S., LUGOSI, G. and MASSART, P. (2000). A sharp concentration inequality with applications. *Random Structures Algorithms* **16** 277–292. . [MR1749290 \(2001m:26072\)](#)
- BOUCHERON, S., LUGOSI, G. and MASSART, P. (2009). On concentration of self-bounding functions. *Electron. J. Probab.* **14** no. 64, 1884–1899. [MR2540852 \(2010k:60058\)](#)
- BURTON, R. M., GOULET, M. and MEESTER, R. (1993). On 1-dependent processes and k -block factors. *Ann. Probab.* **21** 2157–2168. [MR1245304 \(94j:60072\)](#)
- CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal approximation by Stein’s method. Probability and its Applications (New York)*. Springer, Heidelberg. [MR2732624](#)
- CHEN, L. H. Y. and SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. . [MR2073183 \(2005f:60054\)](#)
- DALLAPORTA, S. (2012). Eigenvalue variance bounds for Wigner random matrices. *arXiv preprint arXiv:1203.1597*.
- HOFMANN-CREDNER, K. and STOLZ, M. (2008). Wigner theorems for random matrices with dependent entries: ensembles associated to symmetric spaces and sample covariance matrices. *Electron. Commun. Probab.* **13** 401–414. . [MR2415147 \(2009d:82071\)](#)
- JANSON, S. (2004). Large deviations for sums of partly dependent random variables. *Random Structures Algorithms* **24** 234–248. . [MR2068873 \(2005e:60061\)](#)
- LATAŁA, R. (2005). Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.* **133** 1273–1282 (electronic). . [MR2111932 \(2005i:15041\)](#)
- LUGOSI, G. (2005). *Concentration of Measure Inequalities. Lecture notes*. Unpublished, available at <http://www.econ.upf.edu/~lugosi/anu.pdf>.
- MECKES, M. W. (2004). Concentration of norms and eigenvalues of random matrices. *J. Funct. Anal.* **211** 508–524. . [MR2057479 \(2005c:46109\)](#)
- PAULIN, D. (2012a). Concentration inequalities for Markov chains by Marton couplings. *arXiv preprint*.
- PAULIN, D. (2012b). Concentration of Self-Bounding Functions in Weakly Dependent Spaces by Stein’s Method. *arXiv preprint*.
- SCHENKER, J. H. and SCHULZ-BALDES, H. (2005). Semicircle law and freeness for random matrices with symmetries or correlations. *Math. Res. Lett.* **12** 531–542. [MR2155229 \(2006e:82040\)](#)